Lec 18 MATH 343

Hazard Rate $h(t) := \frac{f(t)}{S(t)}$ which defines a continuous survival distr.

Assume $h(t) = h_0(t) e^{\beta_1 x_1 + \ldots + \beta_p x_p} = h_0(t) e^{\vec{x}\vec{\beta}}$ $\Rightarrow$ each change in x_j implies a multiplicative change in $h(t)$.

We can show

$$f(t) = h_0(t) e^{\vec{x}\vec{\beta}} e^{-\int_0^t h_0(u)\,du\, e^{\vec{x}\vec{\beta}}}$$

Let $t = y$ to match our notation from 342/343

Let $y_1 < y_2 < \ldots < y_n$ be unique values and order $\vec{x}_1, \ldots, \vec{x}_n$ based on that ordering into the rows of X matrix

$$\Rightarrow \mathcal{L}(\vec{\beta}, \underset{\substack{\uparrow \\ \text{nuisance parameter}}}{h_0} ; X, \vec{y}) = \prod_{i=1}^n \underbrace{h_0(y_i)}_{h_i} e^{\vec{x}_i\vec{\beta}} e^{-\overbrace{\int_0^{y_i} h_0(u)\,du}^{H_i}\, e^{\vec{x}_i\vec{\beta}}}$$

add censoring here if necessary

$$\Rightarrow \mathcal{L}(\vec{\beta}, h_1, \ldots, h_n, H_1, \ldots, H_n ; X, \vec{y}) = \prod_{i=1}^n h_i e^{\vec{x}_i\vec{\beta}} e^{-H_i \vec{x}_i \vec{\beta}}$$

Assume: $H_i := \int_0^{y_i} h_0(u)\,du \underset{\substack{\uparrow \\ \text{similar to Riemann integral}}}{\approx} h_0(y_1) + h_0(y_2) + \ldots + h_0(y_i) = \sum_{\ell=1}^{i} h_0(y_\ell) = \sum_{\ell=1}^{i} h_\ell$

$$\mathcal{L}(\vec\beta, \underbrace{h_1,\dots,h_n}_{\text{less nuisances}}; X, \vec y) = \prod_{i=1}^{n} h_i e^{\vec x_i \vec\beta}\, e^{-\left(\sum_{\ell=1}^{i} h_\ell\right) e^{\vec x_i \vec\beta}}$$

$$= \left(\prod_{i=1}^{n} h_i\right) e^{\sum_{i=1}^{n} \vec x_i \vec\beta}\, e^{-\sum_{i=1}^{n}\left(\sum_{\ell=1}^{i} h_\ell\right) e^{\vec x_i \vec\beta}}$$

$$\Rightarrow \ell(\vec\beta, h_1,\dots,h_n; X, \vec y) = \sum_{i=1}^{n} \ln(h_i) + \underbrace{\sum_{i=1}^{n} \vec x_i \vec\beta}_{\vec 1_n^T (X\vec\beta)} - \sum_{i=1}^{n} e^{\vec x_i \vec\beta} \sum_{\ell=1}^{i} h_\ell$$

Let's try to find MLE for h_k where $k \in \{1, 2, \dots, n\}$

$$\frac{\partial}{\partial h_k}[\ell] = \frac{1}{h_k} - \sum_{i=1}^{n} e^{\vec x_i \vec\beta} \sum_{\ell=1}^{i} \frac{\partial}{\partial h_k}[h_\ell] = \frac{1}{h_k} - \sum_{i=1}^{n} e^{\vec x_i \vec\beta} \sum_{\ell=1}^{i} \mathbb{1}_{\ell = k}$$

$$= \frac{1}{h_k} - \sum_{i=1}^{n} e^{\vec x_i \beta}\, \mathbb{1}_{\overbrace{k \le i}^{i \ge k}} = \frac{1}{h_k} - \sum_{i=k}^{n} e^{\vec x_i \vec\beta} \overset{\text{set}}{=} 0 \text{ to find MLE}$$

$$\Rightarrow \frac{1}{h_k} = \sum_{i=k}^{n} e^{\vec x_i \vec\beta} \Rightarrow \hat h_k^{MLE} = \frac{1}{\sum_{i=k}^{n} e^{\vec x_i \vec\beta}}$$

let $c_i := \sum_{\ell=1}^{i} \hat h_\ell^{MLE}$

$$\frac{\partial}{\partial \vec\beta}[\ell] = \frac{\partial}{\partial \vec\beta}\Big[\underbrace{\vec 1_n^T X \vec\beta}_{\vec\beta^T X^T \vec 1_n}\Big] - \sum_{i=1}^{n} \frac{\partial}{\partial \vec\beta}\left[e^{c_i \vec x_i \vec\beta}\right]$$

$$\frac{\partial}{\partial \vec x}\left[e^{\vec a^T \vec x}\right] = \begin{bmatrix} \frac{\partial}{\partial x_1}\left[e^{a_1 x_1} e^{a_2 x_2} \cdots e^{a_n x_n}\right] \\ \vdots \\ \frac{\partial}{\partial x_n}\left[e^{a_1 x_1} e^{a_2 x_2} \cdots e^{a_n x_n}\right] \end{bmatrix} = \begin{bmatrix} a_1 e^{a_1 x_1} \cdots e^{a_n x_n} \\ \vdots \\ a_n e^{a_1 x_1} \cdots e^{a_n x_n} \end{bmatrix} = e^{\vec a^T \vec x} \vec a$$

$$= X^T \vec 1_n - \sum_{i=1}^{n} c_i \vec x_i^T e^{c_i \vec x_i \vec\beta} \overset{\text{set}}{=} \vec 0_p \text{ and solve (no closed form solution).}$$

This can be approximated using an optimizer. And the Fisher Information Matrix can be found as well! $\Rightarrow$ Wald tests for single effects and multiple effects!

Also available to be solved for arbitrary subjects censored!

Causality!

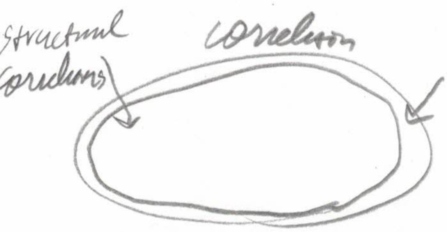
First of all, we learned in 342 that two vectors $\vec{x}, \vec{y} \in \mathbb{R}^n$ have a non-zero dot product. Hence

$$r := \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2} = \frac{(\vec{x} - \bar{x}\vec{1}_n)^T(\vec{y} - \bar{y}\vec{1}_n)}{n s_x s_y}$$

can appear significantly different than zero just by chance. This is called a "spurious correlation". It is not a new concept. $\boxed{\text{Demo}}$

In 340, we called it a Type I error!

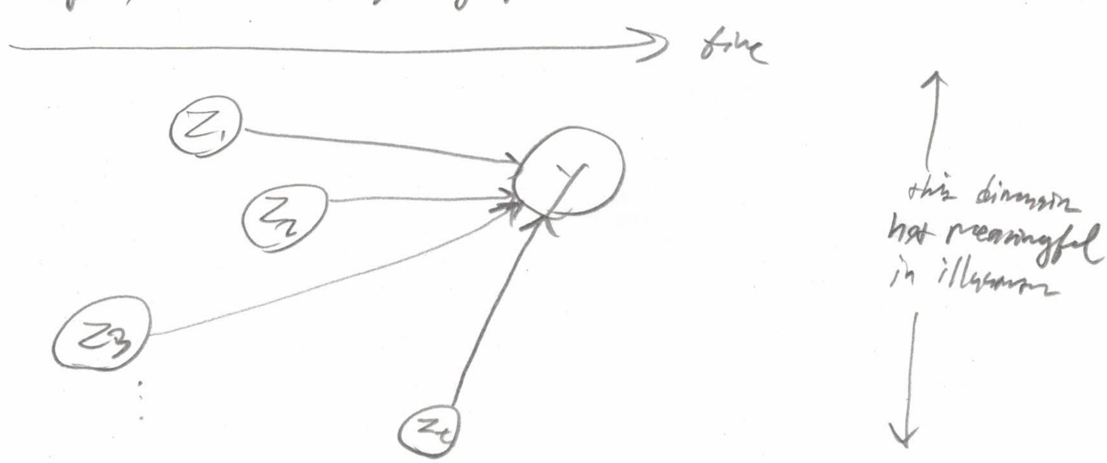
> Spurious correlations are due to chance coincidences and disappear because sort. insig. if n gets larger!



The opposite of spurious is "structural correlation" or just "correlation"; this is based on real, probabilistic dependence. Let's visualize this "structure". First, Assume causes must exist <u>before</u> the phenomenon; this is called the "temporal precedence assumption". Now let's draw a causal diagram, a directed acyclic graph (DAG):

Recall

$y = t(z_1, \ldots, z_t)$



What are the arrows? The arrows mean there is a mathematical relationship between the variables connected where the end of the arrow's variable is the response. E.g.

$y = k_{y,z_1}(z_1) + \eta_{y,z_1}$ ← err, ignorance s.t. $\frac{\partial y}{\partial z_1} \neq 0$ for at least one value of z_1

$y = k_{y,z_2}(z_2) + \eta_{y,z_2}$ ″

⋮

$y = k_{y,z_t}(z_t) + \eta_{y,z_t}$ ″

$y = k_{y,z_i,z_j}(z_i, z_j) + \eta_{y,z_i,z_j}$ $\frac{\partial y}{\partial z_i} \neq 0$ and $\frac{\partial y}{\partial z_j} \neq 0$

⋮

any combination

⋮

$y = t(z_1, \ldots z_t)$ no noise! $\frac{\partial y}{\partial z_i} \neq 0 \ \forall i$

⇒ ~~if you change values on the r.h.s. in the function, then~~ the l.h.s changes for some change in value(s) on the r.h.s.

This is "structural dependence" which is also "causality". Why is it causal? Because if z_1 is "manipulated" by an outside force, due to the structural dependence of k_{y,z_1}.